United States
Department of
Agriculture

Forest Service

**Rocky Mountain
Forest and Range
Experiment Station**

Fort Collins,
Colorado 80526

**Research Paper
RM-309**

Expected Value and Variance of Moran's Bivariate Spatial Autocorrelation Statistic for a Permutation Test

Raymond L. Czaplewski
Robin M. Reich

$$I_{YZ} = \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij} Y_i Z_j}{W\sqrt{m_{Y^2} m_{Z^2}}} \quad [1]$$

$$E[I_{YZ}] = \frac{-m_{YZ}}{(n-1)\sqrt{m_{Y^2} m_{Z^2}}} = \frac{-\rho_{YZ}}{n-1} \quad [25]$$

$$\mathrm{Var}(I_{YZ}) = \left(\frac{\left(\frac{m_{YZ}^2 n}{m_{Y^2} m_{Z^2}}\right)[2(W^2 - S_2 + S_1) + (2S_3 - 2S_5)(n-3) + S_3(n-2)(n-3)] + \left(\frac{-m_{Y^2Z^2}}{m_{Y^2} m_{Z^2}}\right)[6(W^2 - S_2 + S_1) + (4S_1 - 2S_2)(n-3) + S_1(n-2)(n-3)] + n[(W^2 - S_2 + S_1) + (2S_4 - S_6)(n-3) + S_4(n-2)(n-3)]}{(n-1)(n-2)(n-3)W^2} \right) - \left(\frac{m_{YZ}^2}{m_{Y^2} m_{Z^2}}\right)\left(\frac{1}{(n-1)^2}\right) \quad [60]$$

$$S_1 = S_4 + S_3, \quad [39]$$

$$S_2 = 2S_5 + S_6, \quad [40]$$

$$W = 1'W1, \quad [41]$$

$$S_3 = 1'(W \circledast W')1, \quad [42]$$

$$S_4 = 1'(W \circledast W)1, \quad [43]$$

$$S_5 = (W1)'(W'1) = 1'W'W'1 = 1'WW1, \quad [44]$$

$$S_6 = (W1)'(W1) + (W'1)'(W'1) = 1'W'W1 + 1'WW'1 = 1'(W'W + WW')1, \quad [45]$$

Abstract

Moran's *I* statistic has been used by ecologists and geographers alike to test for the presence of spatial autocorrelation in a single variable over a two-dimensional plane. In this paper, we provide the derivation of the expected value and variance of a bivariate version of Moran's *I* for use with multivariate data under the assumption of spatial independence. We also demonstrate that Moran's univariate *I* statistic is a special case of Moran's bivariate I_{YZ}. Results of an extensive Monte Carlo study show that the expected value and variance are reliable for several data sets with moderate sample sizes (n = 40 and 127) and varying degrees of correlation among different bivariate surfaces. For small sample sizes ($n < 40$) at least 5,000 simulations should be used to estimate the P-value to test the null hypothesis of no spatial autocorrelation between two variables.

Acknowledgements

This work was supported by the Forest Inventory and Analysis and the Forest Health Monitoring Programs, USDA Forest Service; and the Forest Group of the Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency. The authors would also like to thank Mei-ching Chen, Jana Anderson, Pat Pellicane, and Doug Brown for the time they spent reviewing drafts of this manuscript.

USDA Forest Service
Research Paper RM-309

April 1993

Expected Value and Variance of Moran's Bivariate Spatial Autocorrelation Statistic for a Permutation Test

Raymond L. Czaplewski
USDA Forest Service
Rocky Mountain Forest and Range Experiment Station[1]

Robin M. Reich
Department of Forest Sciences
Colorado State University

[1]Headquarters is in Fort Collins, CO in cooperation with Colorado State University

Contents

Expected Value and Variance of Moran's Bivariate Spatial Autocorrelation Statistic for a Permutation Test

Raymond L. Czaplewski and Robin M. Reich

Introduction

In analyzing data over large geographical areas, one is often interested in knowing whether or not there is a spatial relationship among the data, and if there is, can one obtain information on the process responsible for generating the observed pattern. Geostatistical techniques, such as Moran's statistic of spatial autocorrelation (Cliff and Ord 1973), spectral analysis (Ripley 1981), and the Mantel test (Mantel 1967), are just a few procedures which permit tests of hypotheses regarding spatial patterns. These tests not only describe the structure of the spatial pattern but they are also capable of detecting the presence of directional components at various scales (Legendre and Fortin 1989). Although these techniques have been applied to univariate data, they can be extended to handle multivariate data as well.

In this paper we provide the derivation of a bivariate version Moran's I, denoted I_{YZ}, for evaluating spatial autocorrelation under the null hypothesis that the magnitude of spatial autocorrelation is no greater than that expected by the chance distribution over space of independent events. We also demonstrate that Moran's univariate I statistic (Cliff and Ord 1973) is a special case of Moran's bivariate I_{YZ}. Finally, we discuss the statistical properties of Moran's bivariate statistic using data from the USDA Forest Service, Southeastern Forest Experiment Station Inventory and Analysis Program, for natural, undisturbed shortleaf pine stands (*Pinus echinata*) in northern Georgia.

Bivariate Moran's I_{YZ}

Define statistic I_{YZ} as an index of spatial autocorrelation between two variables (Y and Z), each of which are observed at n locations (I_{YZ} is analogous to Moran's I for a univariate variable):

$$I_{YZ} = \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij} Y_i Z_j}{W \sqrt{m_{Y^2} m_{Z^2}}} \qquad [1]$$

where

w_{ij} = a scalar that quantifies the degree of spatial association or proximity between locations i and j (e.g., inverse distance between locations i and j, or a 0-1 variable indicating that locations i and j are within some distance range of each other); w_{ij} does not necessarily equal w_{ji}; $w_{ii} = 0$ for all i

Y_i = the observed value of variable Y for plot i, transformed so its mean is zero

$$\frac{\sum_{p=1}^{n} y_p}{n} = \sum_{p=1}^{n} y_p = 0$$

Z_j = the observed value of variable Z for plot j, transformed so its mean is zero

$$\frac{\sum_{p=1}^{n} z_p}{n} = \sum_{p=1}^{n} z_p = 0$$

$$W = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij}$$

$$m_{Y^2} = \frac{\sum_{p=1}^{n} y_p^2}{n} = \mathrm{Var}\,(Y)$$

$$m_{Z^2} = \frac{\sum_{p=1}^{n} z_p^2}{n} = \mathrm{Var}\,(Z)$$

i = $\{1, 2, ..., n\}$
j = $\{1, 2, ..., n\}$
n = number of observed locations, $n \geq 4$.

The denominator in Eq. 1 makes I_{YZ} a dimensionless statistic that can be interpreted as a weighted correlation coefficient in space between variables Y and Z. Thus, one would expect I_{YZ} to range over the interval of -1 to 1. Large positive values of I_{YZ} would indicate a positive spatial autocorrelation between the two response surfaces, while a large negative value would indicate a negative spatial autocorrelation. Cliff and Ord (1981) point out that in general, the upper bound for Moran's univariate I is usually less than 1, although it can exceed 1 for an irregular pattern of weights, w_{ij}, if extreme values are heavily weighted. Thus, depending on the weights used, it is possible for Moran's bivariate I_{YZ} to exceed the limits of -1 and 1.

Assume the bivariate observation $\{y_i, z_i\}$ for location i is a random, spatially independent drawing from one (or separate identical) population(s), and the joint distribution function(s) for Y and Z are unknown. The null hypothesis is that there is no spatial autocorrelation, without any parametric assumptions regarding the joint distribution of Y and Z. Under this null hypothesis, we may consider the set of all $n!$ random permutations, in which the bivariate observations $\{y_k, z_k\}$, $k = \{1, 2, ..., n\}$, are randomly assigned without replacement to location i, $i = \{1, 2, ..., n\}$, regardless of the original spatial location of the kth observation. Under the null hypothesis of no spatial autocorrelation, each bivariate observation is equally likely at any observation location, i.e., the random assignment of any bivariate observation $\{y_k, z_k\}$ to location i is equally likely:

prob(location i has observation $\{y_k, x_k\}$) = $1/n$, for all i and k.

In the following sections, we derive the exact mean and variance of Moran's bivariate I_{YZ} under the null hypothesis for the $n!$ random permutations, and conditional upon the bivariate observations at the n locations. The derivations closely parallel those for Moran's univariate I given by Cliff and Ord (1973) under their randomization assumption. The derivation proceeds in the next sections as follows: various moments of the joint distributions of Y and Z are derived; these expectations are used to compute the expected values of I_{YZ} and I^2_{YZ} (Eqs. 25 and 30); simplifications are introduced to make computations of the expected value of I^2_{YZ} more efficient and less prone to numerical problems (Eq. 58), and then produce an exact equation for the variance of I_{YZ} (Eq. 60); and we then show that the mean and variance of Moran's univariate I is a special case of Moran's bivariate I_{YZ} which we present here (Eqs. 66, 68, and 70). The expected value is denoted by E[•]; the mean is the expected value of I_{YZ}, i.e., $E[I_{YZ}]$; and the variance is denoted $Var(I_{YZ})$ and is equal to $E[I^2_{YZ}] - E[I_{YZ}]^2$ by definition.

Moments of Joint Probability Density Function

Let Y_i be the value of the y-variable for location i, and Z_i be the value of the z-variable for location i during any one of the $n!$ random permutations. Under the null hypothesis, the expected value of the variable Y_i, taken over all possible $n!$ permutations, is by definition:

$$E[Y_i] = y_1 P(Y_i{=}y_1) + y_2 P(Y_i{=}y_2) + ... + y_n P(Y_i{=}y_n) \quad [2]$$

where $P(Y_i{=}y_p)$ denotes the probability that the permutated random variable Y_i, $i \le i \le n$, equals the realized observation y_p, $i \le p \le n$, over all possible $n!$ permutations. There are $(n - 1)!$ permutations among all possible $n!$ permutations in which $Y_i = y_1$, $(n - 1)!$ permutations in which $Y_i = y_2$, and so forth. Therefore, $P(Y_i{=}y_p) = (n - 1)!/n! = 1/n$, and Eq. 2 simplifies to:

$$E[Y_i] = \sum_{p=1}^{n} y_p \left[\frac{(n-1)!}{n!}\right] = \left(\frac{1}{n}\right) \sum_{p=1}^{n} y_p . \quad [3]$$

Since the y values are transformed to center its mean on zero over the n observations without loss of generalization, we can define the $E[Y_i]$ in Eq. 1, to be equal to:

$$E[Y_i] = \frac{\sum_{p=1}^{n} y_p}{n} = 0 . \quad [4]$$

Likewise, the expected values of the variables Y_i and Z_i and their cross-products, taken over all possible $n!$ permutations, are:

$$E[Z_i] = \frac{\sum_{p=1}^{n} z_p}{n} = 0 \quad [5]$$

$$E[Y_i Z_i] = \frac{\sum_{p=1}^{n} y_p z_p}{n} = Cov(YZ) = m_{YZ} \quad [6]$$

$$E[Y_i^2] = \frac{\sum_{p=1}^{n} y_p^2}{n} = Var(Y) = m_{Y^2} \quad [7]$$

$$E[Z_i^2] = \frac{\sum_{p=1}^{n} z_p^2}{n} = Var(Z) = m_{Z^2} \quad [8]$$

$$E[Y_i^2 Z_i^2] = \frac{\sum_{p=1}^{n} y_p^2 z_p^2}{n} = m_{Y^2Z^2} \quad [9]$$

for all i.

In any random permutation under the null hypothesis, a particular location i can have value $Y_i{=}y_k$ with probability $1/n$, and any neighboring unit j ($j{\neq}i$) can take any value $Z_j{=}z_m$ other than $Z_j{=}z_k$ with probability $1/(n-1)$; therefore:

$$E[Y_i Z_j] = \frac{\sum_{i=1}^{n} y_i}{n} \left(\frac{z_1 + ... + z_{i-1} + z_{i+1} + ... z_n}{(n-1)}\right)$$

$$= \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} y_i z_j}{n(n-1)}$$

$$= \frac{\sum_{i=1}^{n} y_i \left(\sum_{p=1}^{n} z_p - z_i\right)}{n(n-1)} \quad [10]$$

The sum of all z_p's in Eq. 10 equals zero by definition (Eqs. 1 and 5), and

$$E[Y_iZ_j] = \frac{-\sum_{i=1}^{n} y_i z_i}{n(n-1)} . \quad [11]$$

Since $\sum_{i=1}^{n} y_i z_i$ in Eq. 11 equals $\sum_{p=1}^{n} y_p z_p$ in Eq. 6,

$$E[Y_iZ_j] = \frac{-nm_{YZ}}{n(n-1)} = \frac{-m_{YZ}}{n-1} \quad [12]$$

where m_{YZ} is defined in Eq. 6. Eq. 12 applies to all possible pairs of locations *i* and *j*, $i \neq j$, in the *n*! random permutations.

In a similar fashion, $E[Y_i^2 Z_j^2]$ can be derived for $i \neq j$ using Eqs. 7, 8, and 9:

$$E[Y_i^2 Z_j^2] = \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} y_i^2 z_j^2}{n(n-1)}$$

$$= \frac{\sum_{i=1}^{n} y_i^2 \left[\left(\sum_{p=1}^{n} z_p^2 \right) - z_i^2 \right]}{n(n-1)}$$

$$= \frac{\sum_{i=1}^{n} y_i^2 \, (nm_{Z^2}) - \sum_{i=1}^{n} y_i^2 \, [z_i^2]}{n(n-1)}$$

$$= \frac{nm_{Y^2} \, (nm_{Z^2}) - nm_{Y^2Z^2}}{n(n-1)} = \frac{nm_{Y^2} m_{Z^2} - m_{Y^2Z^2}}{n-1} \quad [13]$$

where m_{Y^2}, m_{Z^2}, and $m_{Y^2Z^2}$ are defined in Eqs. 7, 8, and 9 respectively. Next, the $E[Y_iZ_jY_jZ_i]$, for $i \neq j$, can be derived in a fashion similar to Eq. 13 using Eqs. 6 and 9:

$$E[Y_iZ_jY_jZ_i] = \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} y_i \, z_i (y_j z_j)}{n(n-1)}$$

$$= \frac{\sum_{i=1}^{n} y_i z_i \left[\left(\sum_{p=1}^{n} y_p z_p \right) - y_i z_i \right]}{n(n-1)}$$

$$= \frac{\sum_{i=1}^{n} y_i z_i \, (nm_{YZ}) - \sum_{i=1}^{n} y_i^2 \, [z_i^2]}{n(n-1)}$$

$$= \frac{nm_{YZ} (nm_{YZ}) - nm_{Y^2Z^2}}{n(n-1)} = \frac{nm_{YZ}^2 - m_{Y^2Z^2}}{n-1} \quad [14]$$

where m_{YZ} and $m_{Y^2Z^2}$ are defined in Eqs. 6 and 9, respectively.

Similarly, the $E[Y_i^2 Z_j Z_k]$, for $i \neq j$, $i \neq k$, and $j \neq k$, is derived as follows using Eqs. 7, 8, and 9:

$$E[Y_i^2 Z_j Z_k] = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\substack{k=1 \\ i \neq j \neq k}}^{n} y_i^2 z_j z_k}{n(n-1)(n-2)}$$

$$= \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} y_i^2 z_j \left[\left(\sum_{p=1}^{n} z_p \right) - z_i - z_j \right]}{n(n-1)(n-2)}$$

$$= \frac{-\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} y_i^2 z_i z_j - \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} y_i^2 z_j^2}{n(n-1)(n-2)}$$

$$= \frac{-\sum_{i=1}^{n} y_i^2 z_i \left[\left(\sum_{p=1}^{n} z_p \right) - z_i \right] - \sum_{i=1}^{n} y_i^2 \left[\left(\sum_{p=1}^{n} z_p^2 \right) - z_i^2 \right]}{n(n-1)(n-2)}$$

$$= \frac{-\sum_{i=1}^{n} y_i^2 z_i [-z_i] - \sum_{i=1}^{n} y_i^2 [nm_{Z^2} - z_i^2]}{n(n-1)(n-2)}$$

$$= \frac{\sum_{i=1}^{n} y_i^2 z_i^2 - nm_{Z^2} \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} y_i^2 z_i^2}{n(n-1)(n-2)}$$

$$= \frac{nm_{Y^2Z^2} - nm_{Z^2} \, nm_{Y^2} + nm_{Y^2Z^2}}{n(n-1)(n-2)}$$

$$= \frac{2m_{Y^2Z^2} - nm_{Y^2} m_{Z^2}}{(n-1)(n-2)} . \quad [15]$$

The $E[Y_iY_mZ_j^2]$, for $i \neq j$, $i \neq m$, and $j \neq m$, is derived similar to Eq. 15:

$$E[Y_iY_mZ_j^2] = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\substack{m=1 \\ i \neq j \neq m}}^{n} y_i \, y_m z_j^2}{n(n-1)(n-2)}$$

$$= \frac{2m_{Y^2Z^2} - nm_{Y^2} m_{Z^2}}{(n-1)(n-2)} . \quad [16]$$

$E[Y_iZ_jY_jZ_k]$, for $i \neq j$, $i \neq k$, and $j \neq k$, is derived similar to Eq. 15 using Eqs. 6 and 9:

$$E[Y_iZ_jY_jZ_k] = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\substack{k=1 \\ i \neq j \neq k}}^{n} y_i y_j z_j z_k}{n(n-1)(n-2)}$$

$$= \frac{\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} y_i y_j z_j \left[\left(\sum_{p=1}^{n} z_p\right) - z_i - z_j\right]}{n(n-1)\,(n-2)}$$

$$= \frac{-\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} y_i y_j z_j z_i - \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} y_i y_j z_j^2}{n(n-1)\,(n-2)}$$

$$= \frac{-\sum_{i=1}^{n} y_i \left[\left(\sum_{p=1}^{n} y_p z_p\right) - y_i z_i\right] z_i - \sum_{j=1}^{n} \left[\left(\sum_{p=1}^{n} y_p\right) - y_j\right] y_j z_j^2}{n(n-1)\,(n-2)}$$

$$= \frac{-\sum_{i=1}^{n} y_i [n m_{YZ} - y_i z_i] z_i - \sum_{j=1}^{n} [-y_j] y_j z_j^2}{n(n-1)\,(n-2)}$$

$$= \frac{-n m_{YZ} \sum_{i=1}^{n} y_i z_i + \sum_{i=1}^{n} y_i^2 z_i^2 + \sum_{j=1}^{n} y_j^2 z_j^2}{n(n-1)\,(n-2)}$$

$$= \frac{-n m_{YZ}\, n m_{YZ} + n m_{Y^2Z^2} + n m_{Y^2Z^2}}{n(n-1)\,(n-2)}$$

$$= \frac{2 m_{Y^2Z^2} - n m_{YZ}^2}{(n-1)\,(n-2)} \, . \qquad [17]$$

$E[Y_i Z_j Y_m Z_i]$ and $E[Y_i Z_j Y_m Z_m]$, for $i \neq j$, $i \neq m$, and $j \neq m$, are derived similar to Eq. 17:

$$E[Y_i Z_j Y_m Z_i] = \frac{\sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j z_i}{n(n-1)\,(n-2)} = \frac{2 m_{Y^2Z^2} - n m_{YZ}^2}{(n-1)\,(n-2)} \qquad [18]$$

and

$$E[Y_i Z_j Y_m Z_m] = \frac{\sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j z_m}{n(n-1)\,(n-2)}$$

$$= \frac{2 m_{Y^2Z^2} - n m_{YZ}^2}{(n-1)\,(n-2)} \, . \qquad [19]$$

Finally, $E[Y_i Z_j Y_m Z_k]$, for $i \neq j$, $m \neq k$, $i \neq m$, $i \neq k$, $j \neq m$, and $j \neq k$, is derived as follows:

$$E[Y_i Z_j Y_m Z_k] = \frac{\sum_{\substack{i=1 \\ i\neq j\neq m\neq k}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{n} y_i y_m z_j z_k}{n(n-1)\,(n-2)\,(n-3)}$$

$$= \frac{\sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j \left[\left(\sum_{p=1}^{n} z_p\right) - z_i - z_j - z_m\right]}{n(n-1)\,(n-2)\,(n-3)}$$

$$= \frac{\sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j [-z_i - z_j - z_m]}{n(n-1)\,(n-2)\,(n-3)}$$

$$= \frac{-\sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j z_i - \sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j^2 - \sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j z_m}{n(n-1)\,(n-2)\,(n-3)} \, . \qquad [20]$$

From Eqs. 18 and 19, $\sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j z_i = \sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j z_m$ in Eq. 20, and

$$E[Y_i Z_j Y_m Z_k] = \frac{-2 \sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j z_i - \sum_{\substack{i=1 \\ i\neq j\neq m}}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n} y_i y_m z_j^2}{n(n-1)\,(n-2)\,(n-3)} \, . \qquad [21]$$

Using the results of Eqs. 16 and 18 and substituting in Eq. 21 we get:

$$E[Y_i Z_j Y_m Z_k] = \frac{-2\left[\frac{2 m_{Y^2Z^2} - n m_{YZ}^2}{(n-1)\,(n-2)}\right] n(n-1)(n-2) - \left[\frac{2 m_{Y^2Z^2} - n m_{Y^2}\, m_{Z^2}}{(n-1)\,(n-2)}\right] n(n-1)(n-2)}{n(n-1)\,(n-2)\,(n-3)}$$

$$= \frac{-2[2 m_{Y^2Z^2} - n m_{YZ}^2] - [2 m_{Y^2Z^2} - n m_{Y^2}\, m_{Z^2}]}{(n-1)\,(n-2)\,(n-3)}$$

$$= \frac{2n m_{YZ}^2 + n m_{Y^2}\, m_{Z^2} - 6 m_{Y^2Z^2}}{(n-1)\,(n-2)\,(n-3)} \, . \qquad [22]$$

First Two Moments of Moran's Bivariate I_{YZ} Statistic

The first two moments of I_{YZ} may be determined under the null hypothesis of no spatial autocorrelation, conditional upon the n observed locations, and using the expected values that were derived in the previous section. The expected value of I_{YZ} in Eq. 1 is determined in Eqs. 23 to 25:

$$E[I_{YZ}] = E\left[\frac{\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}\, Y_i Z_j}{W\sqrt{m_{Y^2}\, m_{Z^2}}}\right] \, . \qquad [23]$$

Since the variables w_{ij}, W, m_{Y^2}, and m_{Z^2} (Eqs. 1, 7, and 8) are known constants, or invariant under random permutations in Eq. 23:

$$E[I_{YZ}] = \frac{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}E[Y_iZ_j]}{W\sqrt{m_{Y^2}\,m_{Z^2}}} \quad . \qquad [24]$$

The $E[Y_iZ_j]$ in Eq. 24 has the same value for all pairs of locations i and j, and using the definition of W in Eq. 1 and the value of $E[Y_iZ_j]$ derived in Eq. 12:

$$E[I_{YZ}] = \frac{\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}\right)\left(\frac{-\,m_{YZ}}{n-1}\right)}{W\sqrt{m_{Y^2}\,m_{Z^2}}}$$

$$= \frac{W\left(\frac{-\,m_{YZ}}{n-1}\right)}{W\sqrt{m_{Y^2}\,m_{Z^2}}} = \frac{-\,m_{YZ}}{(n-1)\sqrt{m_{Y^2}\,m_{Z^2}}} = \frac{-\,\rho_{YZ}}{n-1} \qquad [25]$$

where m_{YZ}, m_{Y^2}, and m_{Z^2} are defined in Eqs. 6, 7, and 8, respectively, and ρ_{YZ} is a measure of the linear correlation between the variables Y and Z.

$E[I^2_{YZ}]$, which is needed to compute the variance of I_{YZ}, is derived from Eq. 1 as follows:

$$E[I^2_{YZ}] = E\left[\left(\frac{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}Y_iZ_j}{W\sqrt{m_{Y^2}\,m_{Z^2}}}\right)^2\right]$$

$$= E\left[\left(\frac{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}Y_iZ_j}{W\sqrt{m_{Y^2}\,m_{Z^2}}}\right)\left(\frac{\sum_{m=1}^{n}\sum_{\substack{k=1\\ m\neq k}}^{n} w_{mk}Y_mZ_k}{W\sqrt{m_{Y^2}\,m_{Z^2}}}\right)\right]$$

$$= \frac{\sum_{\substack{i=1\\ (i\neq j)}}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1\\ (m\neq k)}}^{n}\sum_{k=1}^{n} w_{ij}w_{mk}\,E[Y_iZ_jY_mZ_k]}{W^2 m_{Y^2}\,m_{Z^2}} \quad . \qquad [26]$$

The numerator of Eq. 26 equals:

$$\sum_{\substack{i=1\\ (i\neq j)}}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1\\ (m\neq k)}}^{n}\sum_{k=1}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1\\ i\neq j,\ m\neq k\\ i\neq m,\ j\neq k\\ j\neq m,\ i\neq k}}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1\\ i\neq j,\ m\neq k\\ i\neq m,\ j\neq k\\ (j=m),\ i\neq k}}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1\\ i\neq j,\ m\neq k\\ i\neq m,\ j\neq k\\ j\neq m,\ (i=k)}}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1\\ i\neq j,\ m\neq k\\ (i=m),\ j\neq k\\ j\neq m,\ i\neq k}}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1\\ i\neq j,\ m\neq k\\ i\neq m,\ (j=k)\\ j\neq m,\ i\neq k}}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1\\ i\neq j,\ m\neq k\\ i\neq m,\ j\neq k\\ (j=m),\ (i=k)}}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1\\ i\neq j,\ m\neq k\\ (i=m),\ (j=k)\\ j\neq m,\ i\neq k}}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k] \ . \qquad [27]$$

Since any other summation over i, j, m, and k in Eq. 27 is infeasible for $i\neq j$ and $m\neq k$, Eq. 27 simplifies to the following:

$$\sum_{\substack{i=1\\ (i\neq j)}}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1\\ (m\neq k)}}^{n}\sum_{k=1}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq m\neq k}}^{n} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n} w_{ij}w_{jk}E[Y_iZ_jY_jZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1\\ i\neq j\neq m}}^{n} w_{ij}w_{mi}E[Y_iZ_jY_mZ_i]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n} w_{ij}w_{ik}E[Y_i^2Z_jZ_k]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1\\ i\neq j\neq m}}^{n} w_{ij}w_{mj}E[Y_iY_mZ_j^2]$$

$$+ \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}w_{ji}E[Y_iZ_jY_jZ_i]$$

$$+ \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}w_{ij}E[Y_i^2Z_j^2] \ . \qquad [28]$$

Next, the expectations in Eqs. 13-18, and 22 can be substituted into Eq. 28 to produce the following:

$$\sum_{\substack{i=1 \\ (i\neq j)}}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1 \\ (m\neq k)}}^{n}\sum_{k=1}^{n} w_{ij}w_{mk}\,E[Y_iZ_jY_mZ_k] =$$

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1 \\ i\neq j\neq m\neq k}}^{n} w_{ij}w_{mk}\right)\frac{2nm_{YZ}^2 + nm_{Y^2}m_{Z^2} - 6m_{Y^2Z^2}}{(n-1)(n-2)(n-3)}$$

$$+\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ i\neq j\neq k}}^{n}(w_{ij}w_{jk} + w_{ij}w_{ki})\right)\frac{2m_{Y^2Z^2}-nm_{YZ}^2}{(n-1)(n-2)}$$

$$+\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ i\neq j\neq k}}^{n}(w_{ij}w_{ik} + w_{ij}w_{kj})\right)\frac{2m_{Y^2Z^2}-nm_{Y^2}m_{Z^2}}{(n-1)(n-2)}$$

$$+\left(\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} w_{ij}w_{ji}\right)\frac{nm_{YZ}^2 - m_{Y^2Z^2}}{n-1}$$

$$+\left(\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} w_{ij}^2\right)\frac{nm_{Y^2}m_{Z^2} - m_{Y^2Z^2}}{n-1}\,. \qquad [29]$$

Finally, Eqs. 26 and 29 can be combined to compute $E[I_{YZ}^2]$:

$$E[I_{YZ}^2] = \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{\substack{k=1 \\ i\neq j\neq m\neq k}}^{n} w_{ij}w_{mk}\right)\left(\frac{n[2m_{YZ}^2 + m_{Y^2}m_{Z^2}] - 6m_{Y^2Z^2}}{(n-1)(n-2)(n-3)}\right)}{W^2m_{Y^2}m_{Z^2}}$$

$$+\frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ i\neq j\neq k}}^{n}(w_{ij}w_{jk} + w_{ij}w_{ki})\right)\left(\frac{2m_{Y^2Z^2} - nm_{YZ}^2}{(n-1)(n-2)}\right)}{W^2m_{Y^2}m_{Z^2}}$$

$$+\frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ i\neq j\neq k}}^{n}(w_{ij}w_{ik} + w_{ij}w_{kj})\right)\left(\frac{2m_{Y^2Z^2} - nm_{Y^2}m_{Z^2}}{(n-1)(n-2)}\right)}{W^2m_{Y^2}m_{Z^2}}$$

$$+\left(\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} w_{ij}w_{ji}\right)\left(\frac{nm_{YZ}^2 - m_{Y^2Z^2}}{n-1}\right)\left(\frac{1}{W^2m_{Y^2}m_{Z^2}}\right)$$

$$+\left(\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} w_{ij}^2\right)\left(\frac{nm_{Y^2}m_{Z^2} - m_{Y^2Z^2}}{n-1}\right)\left(\frac{1}{W^2m_{Y^2}m_{Z^2}}\right). \qquad [30]$$

Computational Simplifications

The formula for the $E[I_{YZ}^2]$ in Eq. 30, which is needed to compute the variance of I_{yz}, can be further simplified, thus greatly reducing the number of computations. First, simplifying notation is introduced, and then used in Eq. 30. The notation in Eqs. 31 to 34, 44 is adopted from Cliff and Ord (1973, pp. 25-26):

$$w_{i\cdot} = \sum_{j=1}^{n} w_{ij} \qquad [31]$$

$$w_{\cdot j} = \sum_{i=1}^{n} w_{ij} \qquad [32]$$

$$S_1 = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} w_{ij}(w_{ij} + w_{ji}) = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n}(w_{ij}^2 + w_{ij}w_{ji}) \qquad [33]$$

$$S_2 = \sum_{i=1}^{n}(w_{i\cdot} + w_{\cdot i})^2 = \sum_{i=1}^{n}(w_{i\cdot}^2 + 2w_{i\cdot}w_{\cdot i} + w_{\cdot i}^2)\,. \qquad [34]$$

In addition, we introduce the following notation:

$$S_3 = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} w_{ij}w_{ji} \qquad [35]$$

$$S_4 = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} w_{ij}^2 \qquad [36]$$

$$S_5 = \sum_{i=1}^{n} w_{i\cdot}w_{\cdot i} \qquad [37]$$

$$S_6 = \sum_{i=1}^{n}(w_{i\cdot}^2 + w_{\cdot i}^2)\,. \qquad [38]$$

It can be verified algebraically from Eqs. 35-38 that S_1 and S_2 in Eqs. 33 and 34, 44 equal:

$$S_1 = S_4 + S_3, \qquad [39]$$

$$S_2 = 2S_5 + S_6. \qquad [40]$$

We have found it useful to implement both the univariate and bivariate Moran's I statistics using a computer matrix language. Therefore, the computational formulae for W, S_1, S_2, S_3, S_4, S_5 and S_6 are given in matrix notation:

$$W = \mathbf{1'W1}, \qquad [41]$$

$$S_3 = \mathbf{1'(W \circledast W')1}, \qquad [42]$$

$$S_4 = \mathbf{1'(W \circledast W)1}, \qquad [43]$$

$$S_5 = \mathbf{(W1)'(W'1) = 1'W'W'1 = 1'WW1}, \qquad [44]$$

$$S_6 = \mathbf{(W1)'(W1) + (W'1)'(W'1) = 1'W'W1 + 1'WW'1 = 1'(W'W + WW')1}, \qquad [45]$$

$$S_1 = S_4 + S_3 = \mathbf{1}'(\mathbf{W} \circledast \mathbf{W} + \mathbf{W} \circledast \mathbf{W}')\mathbf{1}, \qquad [46]$$

$$S_2 = 2S_5 + S_6 = \mathbf{1}'(2\mathbf{W}\mathbf{W} + \mathbf{W}'\mathbf{W} + \mathbf{W}\mathbf{W}')\mathbf{1}, \qquad [47]$$

where

- **1** = is a nx1 vector in which all elements are equal to 1
- **W** = is the nxn matrix in which the ijth element equals w_{ij} and all diagonal elements equal zero (i.e., $w_{ii} = 0$ for all i)
- **W′** = is the matrix transpose of **W**
- **A**⊛**B** = denotes element by element multiplication (i.e., the ijth element of **A**⊛ **B** is equal to $a_{ij}b_{ij}$).

The equality in Eq. 46 is derived from Eqs. 33, 35, and 36, and the equality in Eq. 47 is derived from Eqs. 34, 44, 37 and 38.

The first term in Eq. 30 can be simplified to the following using the definition of W in Eq. 1, and Eqs. 31 and 32:

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1\\ i\neq j\neq m\neq k}}^{n}\sum_{k=1}^{n} w_{ij}w_{mk}\right) = \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}\left(\sum_{\substack{m=1\\ m\neq k\neq i\neq j}}^{n}\sum_{k=1}^{n} w_{mk}\right)$$

$$= \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}\left[W - \left(\sum_{\substack{p=1\\ p\neq i}}^{n} w_{ip}\right) - \left(\sum_{\substack{p=1\\ p\neq j}}^{n} w_{jp}\right) - \left(\sum_{\substack{p=1\\ p\neq i\\ p\neq j}}^{n} w_{pi}\right) - \left(\sum_{\substack{p=1\\ p\neq i\\ p\neq j}}^{n} w_{pj}\right)\right]$$

[48]

where w_{ij} is an element of the ith row and jth column of the **W** matrix ($W = \mathbf{1}'\mathbf{W}\mathbf{1}$ Eqs. 41 to 45). In evaluating Eq. 48, w_{ij} and w_{ji} must be subtracted only once from W in Eq. 48. This leads to the following result:

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1\\ i\neq j\neq m\neq k}}^{n}\sum_{k=1}^{n}\right) w_{ij}w_{mk} =$$

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}[W - (w_{i\cdot}) - (w_{j\cdot}) - (w_{\cdot i}-w_{ji}) - (w_{\cdot j}-w_{ij})] \qquad [49]$$

where $w_{i\cdot}$ and $w_{j\cdot}$ in Eq. 49 are the sums of the ith and jth rows of the **W**, and $w_{\cdot i}$ and $w_{\cdot j}$ are the sums of the ith and jth columns of **W**. Next if we use the definition of W in Eq. 1, S_1 in Eq. 33, and S_2 in Eq. 34 we get:

$$\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}\; W\right) - \sum_{i=1}^{n}(w_{i\cdot} + w_{\cdot i})\sum_{j=1}^{n} w_{ij}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{m=1\\ i\neq j\neq m\neq k}}^{n}\sum_{k=1}^{n} w_{ij}w_{mk}\right)$$

$$= -\sum_{j=1}^{n}(w_{j\cdot}+w_{\cdot j})\sum_{i=1}^{n} w_{ij} + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}(w_{ji}+w_{ij})$$

$$= W^2 - \sum_{i=1}^{n}(w_{i\cdot} + w_{\cdot i})\, w_{i\cdot} - \sum_{j=1}^{n}(w_{j\cdot} + w_{\cdot j})w_{\cdot j} + S_1$$

$$= W^2 - \sum_{i=1}^{n}(w_{i\cdot}^2 + w_{\cdot i}w_{i\cdot}) - \sum_{i=1}^{n}(w_{i\cdot}w_{\cdot i} + w_{\cdot i}^2) + S_1$$

$$= W^2 - \sum_{i=1}^{n}(w_{i\cdot}^2 + 2w_{\cdot i}w_{i\cdot} + w_{\cdot i}^2) + S_1$$

$$= W^2 - S_2 + S_1 \; . \qquad [50]$$

This agrees with the derivation by Cliff and Ord (1973, p. 26, Eq. 2.18).

The next term in Eq. 30 is derived as follows:

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n}(w_{ij}w_{jk} + w_{ij}w_{ki})\right) = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n} w_{ij}(w_{jk} + w_{ki})\right)$$

$$= \left[\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}\left(\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n} w_{jk} + \sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n} w_{ki}\right)\right] \; . \qquad [51]$$

Using the notation in Eqs. 31 and 32, Eq. 51 simplifies to:

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n}(w_{ij}w_{jk} + w_{ij}w_{ki})\right)$$

$$= \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}[(w_{j\cdot}-w_{ji}-w_{jj}) + (w_{\cdot i}-w_{ii}-w_{ji})] \; . \qquad [52]$$

Since $w_{ii} = w_{jj} = 0$ by definition in Eq. 1, Eq. 52 reduces to:

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n}(w_{ij}w_{jk} + w_{ij}w_{ki})\right)$$

$$= \left(\sum_{j=1}^{n} w_{j\cdot}\sum_{\substack{i=1\\ i\neq j}}^{n} w_{ij} - \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}w_{ji}\right) + \left(\sum_{i=1}^{n} w_{\cdot i}\sum_{\substack{j=1\\ j\neq i}}^{n} w_{ij} - \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}w_{ji}\right)$$

$$= \sum_{j=1}^{n} w_{j\cdot}(w_{\cdot j}-w_{jj}) + \sum_{i=1}^{n} w_{\cdot i}(w_{i\cdot}-w_{ii}) - 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}w_{ji}$$

$$= 2\sum_{i=1}^{n} w_{i\cdot}w_{\cdot i} - 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij}w_{ji} \; . \qquad [53]$$

From Eqs. 35 and 37, Eq. 53 simplifies to the following term in Eq. 30:

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n}(w_{ij}w_{jk}+w_{ij}w_{ki})\right)=2S_5-2S_3\ . \qquad [54]$$

The next term in Eq. 30 is derived similar to Eqs. 51- 54 using Eqs. 36 and 38 as follows:

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n}(w_{ij}w_{ik}+w_{ij}w_{kj})\right)=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ i\neq j\neq k}}^{n}w_{ij}(w_{ik}+w_{kj})\right)$$

$$=\left[\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}w_{ij}\left(\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n}w_{ik}+\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n}w_{kj}\right)\right]$$

$$=\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}w_{ij}[(w_{i\cdot}-w_{ii}-w_{ij})+(w_{\cdot j}-w_{ij}-w_{jj})]$$

$$=\left(\sum_{i=1}^{n}w_{i\cdot}\sum_{\substack{j=1\\ j\neq i}}^{n}w_{ij}-\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}w_{ij}^2\right)+\left(\sum_{i=1}^{n}w_{\cdot j}\sum_{\substack{i=1\\ i\neq j}}^{n}w_{ij}-\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}w_{ij}^2\right)$$

$$=\sum_{i=1}^{n}w_{i\cdot}(w_{i\cdot}-w_{ii})+\sum_{j=1}^{n}w_{\cdot j}(w_{\cdot j}-w_{jj})-2\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}w_{ij}^2$$

$$=\left(\sum_{i=1}^{n}w_{i\cdot}^2+\sum_{j=1}^{n}w_{\cdot j}^2\right)-2\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}w_{ij}^2$$

$$=S_6-2S_4\ . \qquad [55]$$

Substituting Eqs. 35, 36, 50, 54, and 55 for $E[I_{YZ}^2]$ in Eq. 30:

$$E[I_{YZ}^2]=\frac{\left(\begin{array}{l}(W^2-S_2+S_1)(2nm_{YZ}^2-6m_{Y^2Z^2}+nm_{Y^2}m_{Z^2})\\ +(2S_5-2S_3)(2m_{Y^2Z^2}-nm_{YZ}^2)(n-3)\\ +(S_6-2S_4)(2m_{Y^2Z^2}-nm_{Y^2}m_{Z^2})(n-3)\\ \quad +S_3(nm_{YZ}^2-m_{Y^2Z^2})(n-2)(n-3)\\ \quad +S_4(nm_{Y^2}m_{Z^2}-m_{Y^2Z^2})(n-2)(n-3)\end{array}\right)}{(n-1)(n-2)(n-3)W^2m_{Y^2}m_{Z^2}}\ . \qquad [56]$$

Collecting terms for m_{YZ}^2, $m_{Y^2Z^2}$, and $m_{Y^2}m_{Z^2}$ in Eq. 56:

$$E[I_{YZ}^2]=$$

$$\frac{\left(\begin{array}{l}m_{YZ}^2[(W^2-S_2+S_1)2n-(2S_5-2S_3)n(n-3)+S_3n(n-2)(n-3)]\\ +m_{Y^2Z^2}[-6(W^2-S_2+S_1)+2(2S_5+S_6-2S_3-2S_4)(n-3)-(S_3+S_4)(n-2)(n-3)]\\ +m_{Y^2}m_{Z^2}[(W^2-S_2+S_1)n-(S_6-2S_4)n(n-3)+S_4n(n-2)(n-3)]\end{array}\right)}{(n-1)(n-2)(n-3)W^2m_{Y^2}m_{Z^2}}\ . \qquad [57]$$

Using $S_3+S_4=S_1$ from Eq. 46 and $2S_5+S_6=S_2$ from Eq. 47 and substituting in Eq. 57, and rearranging terms we get:

$$E[I_{YZ}^2]=$$

$$\frac{\left(\begin{array}{l}m_{YZ}^2n[2(W^2-S_2+S_1)-2(S_5-S_3)(n-3)+S_3(n-2)(n-3)]\\ -m_{Y^2Z^2}[6(W^2-S_2+S_1)-2(S_2-2S_1)(n-3)+S_1(n-2)(n-3)]\\ +m_{Y^2}m_{Z^2}n[(W^2-S_2+S_1)-(S_6-2S_4)(n-3)+S_4(n-2)(n-3)]\end{array}\right)}{(n-1)(n-2)(n-3)W^2m_{Y^2}m_{Z^2}}\ . \qquad [58]$$

The variance of I_{YZ} is defined as:

$$\mathrm{Var}(I_{YZ})=E[I_{YZ}^2]-E[I_{YZ}]^2\ . \qquad [59]$$

$E[I_{YZ}]$ in Eq. 59 is already derived in Eq. 24, and $E[I_{YZ}^2]$ is expressed in Eq. 58, therefore:

$$\mathrm{Var}(I_{YZ})=$$

$$\frac{\left(\begin{array}{l}\left(\frac{m_{YZ}^2\,n}{m_{Y^2}m_{Z^2}}\right)[2(W^2-S_2+S_1)+(2S_3-2S_5)(n-3)+S_3(n-2)(n-3)]\\ +\left(\frac{-m_{Y^2Z^2}}{m_{Y^2}m_{Z^2}}\right)[6(W^2-S_2+S_1)+(4S_1-2S_2)(n-3)+S_1(n-2)(n-3)]\\ \qquad +n[(W^2-S_2+S_1)+(2S_4-S_6)(n-3)+S_4(n-2)(n-3)]\end{array}\right)}{(n-1)(n-2)(n-3)W^2}$$

$$-\left(\frac{m_{YZ}^2}{m_{Y^2}m_{Z^2}}\right)\left(\frac{1}{(n-1)^2}\right)\ . \qquad [60]$$

In most applications, the spatial weights are symmetric, i.e., $w_{ij}=w_{ji}$ and $\mathbf{W}=\mathbf{W}'$. In this special case:

$S_1 = 2\,\mathbf{1}'(\mathbf{W}\mathrm{x}\mathbf{W})\mathbf{1}$
$S_2 = 4\,\mathbf{1}'(\mathbf{WW})\mathbf{1}$
$S_3 = S_4 = S_1/2$
$S_5 = S_2/4$
$S_6 = S_2/2$

$$\mathrm{Var}(I_{YZ})=$$

$$\frac{\left(\begin{array}{l}\left(\frac{m_{YZ}^2\,n}{m_{Y^2}m_{Z^2}}\right)\left[2(W^2-S_2+S_1)+(S_1-\frac{S_2}{2})(n-3)+S_1(n-2)(n-3)\right]\\ +\left(\frac{-m_{Y^2Z^2}}{m_{Y^2}m_{Z^2}}\right)[6(W^2-S_2+S_1)+(4S_1-2S_2)(n-3)+S_1(n-2)(n-3)]\\ \qquad +n\left[(W^2-S_2+S_1)+(S_1-\frac{S_2}{2})(n-3)+S_1(n-2)(n-3)\right]\end{array}\right)}{(n-1)(n-2)(n-3)W^2}$$

$$-\left(\frac{m_{YZ}^2}{m_{Y^2}m_{Z^2}}\right)\left(\frac{1}{(n-1)^2}\right)\ . \qquad [61]$$

Univariate Moran's *I* as a Special Case

Moran's bivariate I_{YZ} should include Moran's univariate *I* as a special case, as presented by Cliff and Ord (1973, pp 32-33). The purpose of this section is to support this supposition by demonstrating that the mean and variance of Moran's bivariate I_{YZ} is identical to those of Moran's univariate *I* when the two response variables *Y* and *Z* are identical (e.g. $Y = Z$).

In the univariate case, $y_p = z_p$ for all locations, $1 \le p \le n$. Substituting this identity into Eqs. 6, 7, 8, and 9:

$$m_{YZ} = \frac{\sum_{p=1}^{n} z_p z_p}{n} = \frac{\sum_{p=1}^{n} z_p^2}{n} = \mathrm{Var}(Z) = m_2 \qquad [62]$$

$$m_{Y^2} = \frac{\sum_{p=1}^{n} z_p^2}{n} = \mathrm{Var}(Z) = m_2 \qquad [63]$$

$$m_{Z^2} = \frac{\sum_{p=1}^{n} z_p^2}{n} = \mathrm{Var}(Z) = m_2 \qquad [64]$$

$$m_{Y^2Z^2} = \frac{\sum_{p=1}^{n} z_p^2 z_p^2}{n} = \frac{\sum_{p=1}^{n} z_p^4}{n} = m_4 \qquad [65]$$

where m_2 and m_4 are taken from the notation used by Cliff and Ord (1973, pp. 32-33). Substituting m_2 and m_4 (Eqs. 62-65) in Eq. 24 we get:

$$E[I_{YZ}] = \frac{-m_2}{(n-1)\sqrt{m_2 m_2}} = \frac{-m_2}{(n-1)m_2} = \frac{-1}{(n-1)} \qquad [66]$$

which is identical to the results of Cliff and Ord (1973, p. 32, Eq. 2.31) for the expected value of Moran's univariate *I*.
Substituting m_2 and m_4 (Eqs. 62-65) in Eq. 29:

$$E\left[\left(\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}Y_iZ_j\right)^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{n}{}_{(i\neq j),\,(m\neq k)} w_{ij}w_{mk}E[Y_iZ_jY_mZ_k]$$

$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq m\neq k} w_{ij}w_{mk}\right)\left(\frac{n(2m_2^2+m_2m_2)-6m_4}{(n-1)(n-2)(n-3)}\right)$$

$$+\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ij}w_{jk} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ij}w_{ki}\right)\left(\frac{2m_4-nm_2^2}{(n-1)(n-2)}\right)$$

$$+\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ij}w_{ik} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ij}w_{kj}\right)\left(\frac{2m_4-nm_2m_2}{(n-1)(n-2)}\right)$$

$$+\left(\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}w_{ji}\right)\left(\frac{nm_2^2-m_4}{n-1}\right)$$

$$+\left(\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}^2\right)\left(\frac{nm_2m_2-m_4}{n-1}\right). \qquad [67]$$

By switching the *i* and *j* subscripts in Eq. 67 for $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ij}w_{jk}$ and $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ij}w_{kj}$ we get:

$$E\left[\left(\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}Y_iZ_j\right)^2\right]$$

$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq m\neq k} w_{ij}w_{mk}\right)\frac{3nm_2^2-6m_4}{(n-1)(n-2)(n-3)}$$

$$+\left(\begin{array}{l}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ji}w_{ik} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ij}w_{ki}\\ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ij}w_{ik} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} w_{ji}w_{ki}\end{array}\right)\left(\frac{2m_4-nm_2^2}{(n-1)(n-2)}\right)$$

$$+\left(\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}(w_{ij}+w_{ji})\right)\left(\frac{nm_2^2-m_4}{n-1}\right)$$

and then combining terms:

$$E\left[\left(\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}Y_iZ_j\right)^2\right]$$

$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq m\neq k} w_{ij}w_{mk}\right)\frac{3nm_2^2-6m_4}{(n-1)(n-2)(n-3)}$$

$$+\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{}_{i\neq j\neq k} (w_{ij}+w_{ji})(w_{ik}+w_{ki})\right)\left(\frac{2m_4-nm_2^2}{(n-1)(n-2)}\right)$$

$$+\left(\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}(w_{ij}+w_{ji})\right)\frac{nm_2^2-m_4}{n-1} \qquad [68]$$

which is identical to the results given by Cliff and Ord (1973, p. 33, Eq. 2.33) for $E[I_{YZ}^2]$.

The next step is to verify that $E[I_{YZ}^2]$ in Eq. 58 agrees with the equation for the same quantity in Cliff and Ord (1973, p. 33, Eq. 2.34). Substituting m_2 and m_4 (Eqs. 62-65) into Eq. 58:

$$E[I_{YZ}^2] = \frac{\left(\begin{array}{c} n[2(W^2-S_2+S_1) + (2S_3-2S_5)(n-3) + S_3(n-2)(n-3)] \\ + \left(\frac{-m_{Y^2Z^2}}{m_{Y^2}m_{Z^2}}\right)[6(W^2-S_2+S_1) + (4S_1-2S_2)(n-3) + S_1(n-2)(n-3)] \\ + n[(W^2-S_2+S_1) + (2S_4-S_6)(n-3) + S_4(n-2)(n-3)] \end{array}\right)}{(n-1)(n-2)(n-3)W^2}$$

$$= \frac{\left(\begin{array}{c} n\left[\begin{array}{c} 3(W^2-S_2+S_1) + (2S_3+2S_4-2S_5-S_6)(n-3) \\ + (S_3+S_4)(n-2)(n-3) \end{array}\right] \\ + \left(\frac{-m_{Y^2Z^2}}{m_{Y^2}m_{Z^2}}\right)[6(W^2-S_2+S_1) + (4S_1-2S_2)(n-3) + S_1(n-2)(n-3)] \end{array}\right)}{(n-1)(n-2)(n-3)W^2}. \quad [69]$$

Since $S_3 + S_4 = S_1$ from Eq. 46 and $2S_5 + S_6 = S_2$ from Eq. 47, Eq. 69 simplifies to:

$$E[I_{YZ}^2] = \frac{\left(\begin{array}{c} n[3(W^2-S_2+S_1) + (2S_1-S_2)(n-3) + S_1(n-2)(n-3)] \\ + \left(\frac{-m_{Y^2Z^2}}{m_{Y^2}m_{Z^2}}\right)[6(W^2-S_2+S_1) + (4S_1-2S_2)(n-3) + S_1(n-2)(n-3)] \end{array}\right)}{(n-1)(n-2)(n-3)W^2}$$

$$= \frac{\left(\begin{array}{c} n[3W^2 + (n^2-3n+3)S_1 - nS_2] \\ + \left(\frac{-m_{Y^2Z^2}}{m_{Y^2}m_{Z^2}}\right)[6W^2 + (n^2-n)S_1 - 2nS_2] \end{array}\right)}{(n-1)(n-2)(n-3)W^2}. \quad [70]$$

Eq. 70 is identical to Eq. 2.34 in Cliff and Ord (1973). This demonstrates that the variance of Moran's bivariate I_{YZ} in Eq. 60 equals the variance of Moran's univariate I under the assumptions in Eqs. 62-65.

Statistical Properties of Moran's Bivariate I_{YZ} DATA

The statistical properties of Moran's bivariate I_{YZ} were evaluated using a subset of USDA Forest Service inventory data from Bechtold et al. (1991). The data were originally used to detect changes in net basal area growth of natural, undisturbed shortleaf pine stands in Georgia between 1961-72 and 1972-82. Because of high disturbance rates the number of sample plots used to estimate growth varied in each period (Bechtold et al. 1991); 127 plots for the 1961-72 period and 40 plots for the 1972-82 period.

Data used in the analysis included the natural logarithm of gross annual pine basal area growth per acre (G), site index (S) (base age 50), natural logarithm of stand age (A) (midpoint of 10-year class), natural logarithm of the number of trees per acre (N), ratio of pine basal area to basal area of all species (P), and natural logarithm of annual basal area mortality per acre (M). For a more detailed description of the data see Bechtold et al. (1991).

Simulation

Permutation procedures were used to estimate the mean, variance, skewness, and P-value of Moran's bivariate I_{YZ} under the null hypothesis of no spatial autocorrelation between two variables. Since the number of possible permutations tends to be large for even small data sets, an approximation to the permutation procedure was used to obtain estimates of the population parameters under the null hypothesis. This was accomplished by randomly assigning the response variables to the n distinct geographical locations and calculating I_{YZ}. This process was repeated 200,000 times with replacement for each of the 15 pairs of variables associated with the two growth cycles. In calculating I_{YZ} the inverse distance between plots was used as a weighting factor.

The simulations were used to detect any errors in the derivation of I_{YZ} (Eq. 25) and its variance (Eq. 60). A percent difference for I_{YZ} was computed as the difference between the average of the 200,000 estimates of I_{YZ} from the simulations and the expected value (Eq. 25) under the null hypothesis, divided by the expected value. This was done for each of the 15 pairs of variables in each growth period. In addition, the variance of the mean (i.e. simulation variance) was also computed as the variance among the 200,000 estimates of I_{YZ}. In general, the variance of the mean is the best estimate of the variability and can be used to evaluate whether the derivation of the variance of I_{YZ} given by Eq. 60 is correct.

Finally, the P-value for the realized value of I_{YZ} (say I_0) was calculated as the proportion of the I_{YZ}'s from the 200,000 simulations that were more extreme than I_0, or equivalently, the probability under the null hypothesis given by $P(I_{YZ} > I_0)$. A P-value was also calculated using the normal and Pearson type III distributions.

Comparison of the Expected Value and Simulation Mean

The expected value of I_{YZ} (Eq. 25) agreed with the simulation mean of Moran's bivariate I_{YZ}. In general, as the number of observations increased from 40 (second growth cycle) to 127 (first growth cycle) the magnitude of the percent difference in I_{YZ} decreased from 2.7 percent to -0.10 percent. The largest percent difference was observed when the linear correlation between any two variables was near zero (e.g., $\rho_{YZ}\approx 0$). Since the expected value of Moran's bivariate test statistic I_{YZ} approaches zero as the linear correlation between any two variables approaches zero, the percent difference tends toward infinity. Thus, this trend is likely a numerical artifact and not an indication of an error in the derivation of the expected value of Moran's bivariate test statistic.

For small sample sizes (n = 40) the simulation mean tends to underestimate the expected value with increasing magnitude as the linear correlation between any two variables increased in a positive direction. This trend was not significant, however, at the 0.05 level of significance. As the sample size increased (n = 127) the percent difference was observed to be independent of the linear correlation between any two variables. As the sample size increases, we expect the percent difference to approach zero.

Ratio of Variances

The ratios of the simulation variance to the variance of Moran's bivariate I_{YZ} under the null hypothesis were all less than 1. A ratio less than 1 indicates an underestimation of the true variance, while a ratio greater than 1 would indicate an overestimation. As the sample size increased from 40 to 127 the average ratio increased from 0.961 to 0.985, respectively. We conjecture that as the sample size increases the simulation variance will approach the true value given by Eq. 60.

Skewness

The skewness for Moran's bivariate I_{YZ} is given in Table 1. Except for two cases (G-N and S-P) the absolute value of the skewness was less than 0.01 for both data sets analyzed in this study, indicating that Moran's bivariate I_{YZ} tends to be normally distributed for sample sizes $n > 40$. Mielke (1986) points out that if the absolute value of the skewness is less than 0.01, one can reliably assume a normal distribution in estimating P-values. In the case when the absolute value of the skewness is greater than 0.01, the Pearson type III distribution can be used to account for skewness in calculating P-values.

Table 1. Skewness of Moran's bivariate I_{YZ} associated with testing the null hypothesis of no spatial autocorrelation between two variables.

Variables[1/]	Data set 1 n = 127	Data set 2 n = 40
G-S	0.0026	0.0062
G-A	-0.0037	0.0004
G-N	0.0055	0.0159[2]
G-P	0.0046	0.0082
G-M	-0.0014	-0.0017
S-A	-0.0001	0.0007
S-N	0.0010	-0.0071
S-P	0.0028	0.0119[2]
S-M	-0.0007	0.0084
A-N	-0.0013	0.0034
A-P	-0.0013	0.0030
A-M	0.0013	-0.0018
N-P	0.0034	0.0069
N-M	0.0047	0.0088
P-M	0.0002	0.0058

1/ *See the text for definitions of the variables.*

2/ *If the absolute value of the skewness is greater than 0.01, one cannot reliably assume a normal distribution in estimating P-values (Mielke 1986).*

P-value

Table 2 compares the P-values for Moran's bivariate I_{YZ} under the null hypothesis of no spatial autocorrelation between two variables. P-values were calculated using the normal distribution, Pearson's type III distribution which takes into consideration the amount of skewness (Table 1), and as the proportion of the 200,000 simulated I_{YZ}'s that were more extreme than the realized value of I_{YZ}. Since the absolute value of the skewness for Moran's bivariate I_{YZ} was less than 0.01 for all but two simulations, P-values based on the normal and Pearson type III distribution did not differ appreciably from one another. Comparing these P-values with those obtained from the simulation indicate that both the normal and Pearson's type III distribution provide a reasonable approximation of the tail probabilities under the null hypothesis of no spatial autocorrelation.

Number of Simulations When Using Small Sample Sizes

The results of this study demonstrate the asymptotic normality of Moran's bivariate I_{YZ} under the null hypothesis for two data sets with sample sizes $n \geq 40$. In this case, it is sufficient to take $z=(I_{YZ}-E[I_{YZ}])/V(I_{YZ})^{1/2}$ as a standard normal variate. The null hypothesis of no spatial autocorrelation between two variables defined by Y_i and Z_i is rejected if $|z| > z_{\alpha/2}$. If I_{YZ} is significantly larger than $E[I_{YZ}]$, this would indicate a positive spatial autocorrelation between the two variables, while values significantly less than $E[I_{YZ}]$ would indicate a negative spatial autocorrelation between the two variables.

Table 2. Comparison of P-values for a 1-tailed test of the null hypothesis of no spatial autocorrelation between two variables.

Variables[1/]	Normal[2/]	Simulation[3/]	Pearson[4/]
	Data set 1, n = 127		
G-S	.499	.487	.499
G-A	.313	.317	.311
G-N	.265	.270	.264
G-P	.293	.273	.292
G-M	.350	.348	.348
S-A	$4.7x10^{-5}$	$7.0x10^{-5}$	$0.7x10^{-5}$
S-N	.204	.198	.203
S-P	.099	.097	.097
S-M	.411	.404	.412
A-N	.265	.260	.263
A-P	.089	.083	.088
A-M	.092	.088	.090
N-P	.439	.421	.439
N-M	.115	.102	.114
P-M	.156	.149	.155
	Data set 2, n = 40		
G-S	$7.8x10^{-4}$	$30.x10^{-4}$	$6.1x10^{-4}$
G-A	.023	.024	.020
G-N	$4.7x10^{-4}$	$30.x10^{-4}$	$.2x10^{-4}$
G-P	$8.5x10^{-5}$	$90.x10^{-5}$	$7.2x10^{-5}$
G-M	.292	.318	.325
S-A	.052	.046	.048
S-N	.086	.075	.083
S-P	.100	.098	.098
S-M	.382	.383	.379
A-N	.075	.065	.071
A-P	.066	.059	.063
A-M	.317	.298	.312
N-P	.008	.011	.007
N-M	.116	.098	.110
P-M	.299	.294	.294

1/ *See the text for the definitions of the variables.*

2/ *Calculated using the expected value and variance of Moran's Bivariate I_{YZ} under the null hypothesis of no spatial autocorrelation.*

3/ *Based on 200,000 Monte Carlo simulations.*

4/ *Based on the Pearson type III distribution using the mean, variance and skewness from a Monte Carlo simulation.*

Cliff and Ord (1981) point out that for small sample sizes ($n < 50$) the assumption of normality may not be valid. The departure from normality may be due to several factors: 1) the spatial distribution of the observations within the study area; 2) the weights used (w_{ij}); 3) the distribution of the variables Y_i and Z_i; and 4) the sample size, n (Cliff and Ord 1981). Thus, for small sample sizes we recommend using a Monte Carlo simulation to approximate the empirical distribution of Moran's bivariate I_{YZ} under the null hypothesis.

To estimate the mean and skewness we recommend at least 150,000 simulations, and at least 200,000 simulations for estimating the variance. This is based on

B

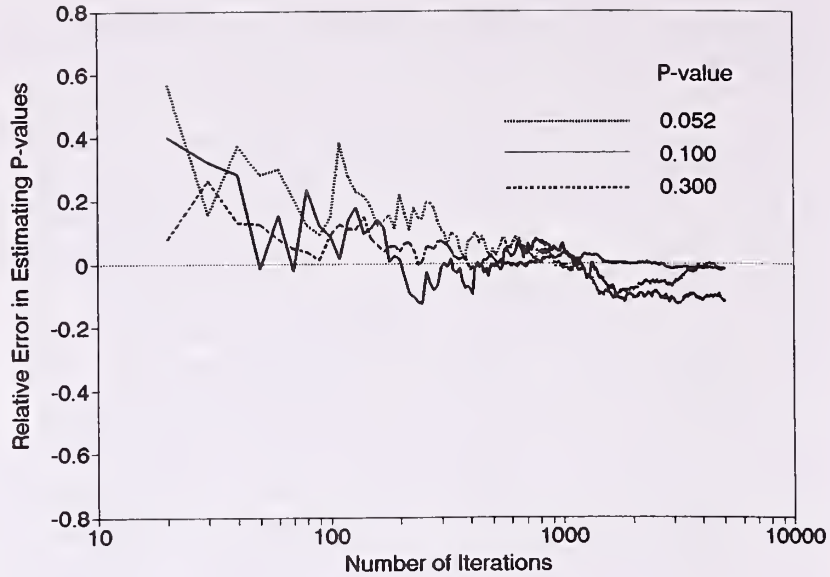

Figure 1. Relative error associated with estimating a range of P-values under the null hypothesis of no spatial autocorrelation as a function of the number of simulations and a sample size n = 40. Figure 1a depicts the P-values 0.0008, 0.008, and 0.023; while Figure 1b depicts the P-values 0.052, 0.10, and 0.30.

evaluations of smaller numbers of simulations during the progression of this study. In spite of the large number of simulations required to estimate the mean, variance, and skewness under the null hypothesis, stable estimates of the P-value can be obtained using a smaller number of simulations. Figure 1 gives examples of the relative error associated with estimating a range of P-values under the null hypothesis of no spatial autocorrelation as a function of the number of simulations and a sample size of n = 40. For P-values greater than 0.05 reasonable estimates can be obtained with 5,000 simulations while P-values less than 0.05 require at least 10,000 simulation (not shown).

Summary

In this paper, we derive the expected value and variance of a bivariate version of Moran's I, a measure of spatial autocorrelation. Results from a Monte Carlo study indicate that for moderate to large sample sizes (n > 40) Moran's bivariate I_{YZ} tends to be normally distributed. The assumption of normality begins to deteriorate for sample sizes less than 40.

For small sample sizes (n < 40), a better approximation to the distribution of the test statistic is needed. One possible approach is to derive the third and possibly the fourth moments of the statistic. This has not been pursued to date because of the complexity of evaluating the higher order moments under the assumption of randomization. Until the problem is solved, the use of Monte Carlo simulations for inference with small sample sizes would be preferable to that of the normal approximation.

Literature Cited

Bechtold, W.A.; Ruark, G.A.; Lloyd, F.T. 1991. Changing stand structure and regional growth reductions in Georgia's natural pine stands. Forest Science. 37: 703-717.

Cliff, A.D.; Ord, J.K. 1973. Spatial autocorrelation. London: Pion Limited. 178 p.

Cliff, A.D.; Ord, J.K. 1981. Spatial processes. London: Pion Limited. 263 p.

Legendre, P.; Fortin, A.J. 1989. Spatial pattern and ecological analysis. Vegetatio. 80: 107-138.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research. 27: 209-220.

Mielke, P.W., Jr. 1986. Nonparametric statistical analysis: some metric alternatives. Journal Statistical Planning Inference. 13: 377-387.

Ripley, B.D. 1981. Spatial statistics. New York: John Wiley & Sons. 252 p.

Czaplewski, R.L; Reich, R.M. 1993. Expected value and variance of Moran's bivariate spatial autocorrelation statistic for a permutation test. Res. Pap. RM-309. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 13 p.

Moran's I statistic has been used by ecologists and geographers alike to test for the presence of spatial autocorrelation in a single variable over a two-dimensional plane. In this paper, we provide the derivation of the expected value and variance of a bivariate version of Moran's I for use with multivariate data under the assumption of spatial independence. We also demonstrate that Moran's univariate I statistic is a special case of Moran's bivariate I_{YZ}. Results of an extensive Monte Carlo study show that the expected value and variance are reliable for several data sets with moderate sample sizes ($n = 40$ and 127) and varying degrees of correlation among different bivariate surfaces. For small sample sizes ($n < 40$) at least 5,000 simulations should be used to estimate the P-value to test the null hypothesis of no spatial autocorrelation between two variables.

Keywords: Permutations, shortleaf, spatial autocorrelation.

Rocky
Mountains

Southwest

Great
Plains

U.S. Department of Agriculture
Forest Service

Rocky Mountain Forest and Range Experiment Station

The Rocky Mountain Station is one of eight regional experiment stations, plus the Forest Products Laboratory and the Washington Office Staff, that make up the Forest Service research organization.

RESEARCH FOCUS

Research programs at the Rocky Mountain Station are coordinated with area universities and with other institutions. Many studies are conducted on a cooperative basis to accelerate solutions to problems involving range, water, wildlife and fish habitat, human and community development, timber, recreation, protection, and multiresource evaluation.

RESEARCH LOCATIONS

Research Work Units of the Rocky Mountain Station are operated in cooperation with universities in the following cities:

Albuquerque, New Mexico
Flagstaff, Arizona
Fort Collins, Colorado*
Laramie, Wyoming
Lincoln, Nebraska
Rapid City, South Dakota

*Station Headquarters: 240 W. Prospect Rd., Fort Collins, CO 80526